Hematology/oncology Clinics of North America
Volume 14 • Number 4 • August 2000

UNDERSTANDING CLINICAL TRIALS

INTERMEDIATE MARKERS AS SURROGATE ENDPOINTS IN CANCER RESEARCH

Arthur Schatzkin MD, DrPH

Nutritional Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

Address reprint requests to
Arthur **Schatzkin**, MD, DrPH
National Cancer Institute
6120 Executive Boulevard, EPS 7032
Bethesda, MD 20892-7232

This article addresses theoretical and practical problems in using surrogate endpoints in experimental and observational studies of cancer.

DEFINITION OF SURROGATE ENDPOINT

In an intervention study (clinical trial), a surrogate endpoint for incident cancer yields a valid test of the null hypothesis of no association between the intervention (treatment) and cancer. [27] In other words, the effect of an intervention on the surrogate is concordant with its effect on cancer incidence. For an observational epidemiologic study, the association of an exposure with the surrogate parallels its association with cancer incidence. (*Concordant* and *parallel* imply proportionality, whereby a large change in the surrogate endpoint implies a large change in cancer incidence and a small change in the surrogate means a small change in cancer incidence.) If a putative surrogate endpoint meets these conditions, it can be considered a valid surrogate for that cancer.

ATTRACTION OF STUDIES USING SURROGATE ENDPOINTS

The enthusiasm for surrogate endpoints derives from the relative rarity of cancer in the general population. The age-adjusted annual incidence of breast cancer among women in the United States, for

example, is about 100 per 100,000, or 0.1%. The incidence of colorectal cancer among men and women combined is around 50 per 100,000, or only 0.05%. [29A] Thus, although cancer is a leading cause of morbidity and mortality in developed countries, even the most common malignancies occur relatively infrequently.

The implications of this simple fact for medical research are straightforward: intervention or prospective observational epidemiologic studies with incident cancer endpoints must be large, long, and costly. Such studies must yield hundreds, or even thousands, of cancers to have adequate statistical power to detect a meaningful treatment effect (intervention study) or relative risk (epidemiologic investigation). The ongoing Women's Health Initiative, for example, requires several tens of thousands of participants to have adequate power to detect reasonable reductions in the incidence of breast and colorectal cancer. [39] (Even case-control studies today have become fairly costly, multiyear efforts.) Studies with surrogate endpoints are attractive because they are potentially smaller, shorter, and much less expensive than their counterparts with cancer endpoints.

RANGE OF POTENTIAL SURROGATE ENDPOINTS FOR CANCER

A host of biologic phenomena--biomarkers in the most general sense--could potentially serve as cancer surrogates. With the explosion of knowledge concerning molecular and cell biology, this list is growing. Potential surrogate biomarkers can be categorized as follows:

1. *Alterations in the microscopic or gross characteristics of tissues.* Pre-neoplastic or frankly neoplastic changes are obvious candidates for surrogate endpoints. Examples include colorectal adenomatous polyps (as surrogates for colorectal cancer), [31] cervical intraepithelial neoplasia (for squamous cell carcinoma of the cervix), [25] bronchial metaplasia (a possible preneoplastic state for lung cancer), [24] and dysplastic changes in the esophagus (for esophageal cancer). [9]
2. *Imaging techniques for detecting histologic change.* Examples of indicative imaging techniques include mammographic parenchymal patterns as a surrogate for breast carcinogenesis, [30] and ovarian ultrasound abnormalities in ovarian cancer. [21]
3. *Cellular phenomena.* Surrogate cellular phenomena include several indices of epithelial cell proliferation, including tritiated thymidine or bromodeoxyuridine incorporation into DNA, proliferating cell nuclear antigen (PCNA), and Ki67. [3] Measures of apoptosis [4] have recently been proposed as potential surrogate endpoints, as has the ratio of proliferation to apoptosis. In AIDS research, CD4 cell counts have been used as surrogates for critical AIDS endpoints. [37]
4. *Molecular markers.* A plethora of potential molecular surrogates have been suggested. Examples include specific somatic mutations in cancer-related genes (such as *ras* or *p53*), both DNA hypo- and hypermethylation of specific genes at various anatomic sites, and gene expression products. [8] [12] Chemical DNA adducts should be considered, not as indicators of exposure (which they might well be), but as markers of an integrated metabolic process that is further removed from the exposure itself. [17]
5. *Infection.* Infectious processes have been implicated in a number of cancers, and these infections could be viewed as surrogate endpoints. Examples include infections with human papillomavirus (HPV) in cervical carcinogenesis, [33] *Helicobacter pylori* in gastric cancer, [26] and human T-cell lymphotrophic virus-1 (HTLV1) in adult T-cell leukemia. [6]
6. *Bioactive substances in blood and tissue.* Examples of bioactive substances include blood and tissue estrogens or androgens, various antioxidants (again, both in blood and in specific tissues), and growth factors. This category of potential surrogates is particularly interesting because the marker--blood estrogen levels, [10] for example--may not be found directly in the target tissue but

may still properly be considered a potential surrogate endpoint--in this case, for breast cancer.

A central question is how to determine whether a particular marker is really a valid surrogate for cancer. Much of this article discusses the logical considerations and empirical data pertaining to this question.

CAUSAL MODELS OF CANCER SURROGACY

The logic underlying the evaluation of surrogate validity is well illustrated by a series of causal pathway models. The simplest causal pathway involving a potential surrogate is depicted in Figure 1 . Here, E1 represents some exposure, an environmental or host factor. (E1 could be anything from a chemopreventive agent to a deleterious risk factor.) According to this simplified model, a change in E1 necessarily alters surrogate endpoint (S) positivity which in turn modifies the likelihood of incident cancer. S, by this definition, is a valid surrogate for cancer.

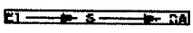

Figure 1. A simple causal pathway model involving a potential surrogate. E1 represents some exposure, an environmental or host factor, such as a chemopreventive agent or deleterious risk factor. A surrogate endpoint (S) is altered by a change in E1 and that in turn modifies the likelihood of incident cancer (CA).

The scenario shown in Figure 1 rarely occurs. Far more realistic are the situations reflected in Figures 2 and 3 . In Figure 2 , E1 modulates carcinogenesis through two alternative pathways, one through S, the other through another marker, M2. To the extent that E1 operates through the alternative M2 pathway--which means that S is not a necessary component of the cancer-causing process--it cannot be assured that S is a valid surrogate in studies involving E1. The reason for this lack of certainty is that E1 might influence M2 in a way that offsets its effect on S, the final effect on cancer simply being unknown. If E1, for example, were to increase M2 positivity, E1 could actually end up *increasing* cancer incidence, even while E1 was reducing S positivity (and thereby giving at least a superficial impression of being anticarcinogenic).

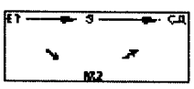

Figure 2. A more realistic situation than is represented in Figure 1 . E1 modulates carcinogenesis through two alternative pathways, one through S and the other through another marker (M2). Because of the alternative M2 pathway, investigators cannot be sure that S is a valid surrogate in studies involving E1. E1 might influence M2 in a way that offsets its effect on S and the final effect on cancer is unknown.

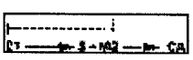

Figure 3. The joint action of two markers (S and M2) is necessary for the development of cancer. E1 may affect either S or M2, but investigators cannot be sure that S is a valid cancer surrogate in studies of exposure to E1 because that exposure may affect S and M2 in counterbalancing ways.

In Figure 3 , the joint action of two markers, S and M2, is necessary for the development of cancer. E1 may affect either S or M2. Again, S cannot be established as a valid cancer surrogate in studies of exposure E1, because that exposure may affect S and M2 in counterbalancing ways.

The scenarios in Figures 2 and 3 , although qualitatively more complex than that represented in Figure 1 , are nevertheless themselves simplified and idealized. Given the complex, multilevel cascade of events that underlie cell growth and inhibition, one can easily envision still more elaborate combinations of pathways reflected in Figures 2 and 3 .

Although the emphasis here has been on etiologic and intervention studies of incident cancer, this

discussion also applies to treatment research, in which one is interested not in the occurrence of a malignancy but in the factors contributing to or preventing mortality from a given cancer.

HYPERPROLIFERATION

An extensive literature addresses problems of proliferation biology. [38] (Although here the focus is primarily on colorectal epithelial cell proliferation, the arguments presented apply to hyperproliferation in a variety of tissues.) A number of cell proliferation assays have been developed; these assays have been suggested as potential surrogates for cancer because the dysregulation of cell growth characterizes malignancy. Figure 4 depicts causal events that may be involved in the relation between hyperproliferation and the neoplastic process. Focusing only on the upper portion of this diagram, a single pathway leads from normal epithelium to hyperproliferative epithelium to neoplasia and cancer. This pathway implicitly underlies the use of hyperproliferation as a valid surrogate for cancer.

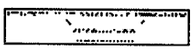

Figure 4. Causal events potentially involved in the relation between hyperproliferation and the neoplastic process. In the upper portion of the figure a single pathway going from normal epithelium to hyperproliferative epithelium to neoplasia/cancer can be seen. This pathway implicitly underlies using hyperproliferation as a valid surrogate for cancer. Hyperproliferation may not be necessary for colorectal carcinogenesis, however, which renders uncertain its validity as a surrogate endpoint.

As the rest of the figure illustrates, however, hyperproliferation may not be necessary for colorectal carcinogenesis. An alternative pathway to neoplasia and cancer may bypasses hyperproliferation. The problem is that the effect of an intervention agent E1 on this alternative pathway is unknown and may in fact counterbalance the effect through the hyperproliferation pathway. Two scenarios here are revealing: (1) the agent E1 reduces proliferation but at the same time reduces apoptosis and therefore has no effect on colorectal cancer; or (2) the agent has no effect on proliferation but does increase apoptosis, thereby reducing colorectal cancer incidence. In both cases, a hyperproliferation assay gives the wrong answer about an intervention's effect on colorectal cancer; by definition, hyperproliferation would not be a valid surrogate for cancer in studies of E1.

It is important to emphasize that the proliferation marker does not necessarily give the wrong answer about the agent's effect on cancer; the proliferation data may, in fact, give the right answer. The problem is the uncertainty that flows from the existence of multiple alternative pathways to cancer. Given this a priori theoretical uncertainty for a putative marker such as hyperproliferation, how can the validity of such a marker be evaluated?

EVALUATING POTENTIAL SURROGATE ENDPOINTS

The answer is to integrate the marker, in this case, epithelial cell proliferation, in observational epidemiologic studies or clinical trials that have colorectal cancer or neoplasia as an endpoint. Such integration can elucidate the causal structure underlying the relations of interventions (or exposures), potential surrogate endpoints, and cancer. The specific data for revealing this underlying structure come from investigating three questions:

1. Is the potential surrogate endpoint associated with cancer?
2. Is the intervention (or exposure) associated with the potential surrogate?

3. Does the potential surrogate endpoint mediate the relation of the intervention (or exposure) to cancer?

Standard epidemiologic measures are pertinent to answering these questions. These measures can be defined with reference to the 2 x 2 table shown in Figure 5 . (For simplicity, surrogate endpoints here are either positive or negative. The arguments presented, however, are germane to markers measured as continuous variables.) For a cohort study, the association between the surrogate endpoint and cancer would be reflected in the relative risk (RR), defined as [a/(a+b)]/[c/(c+d)]. For a case-control study, the RR would be estimated by the odds ratio, defined as ad/bc. A relative risk (or odds ratio) of 1.0 indicates no association between the potential surrogate endpoint and cancer. The attributable proportion (AP) is an epidemiologic measure that indicates the proportion of cancer that is attributable to surrogate endpoint positivity. AP=S(1-1/R) where R = RR and S=sensitivity, defined as a/(a+c). An AP of 1.0 means that marker positivity is necessary for the development of cancer; that is, the carcinogenic pathway must go through this positive marker or, to put it another way, there is no alternative pathway bypassing this marker.

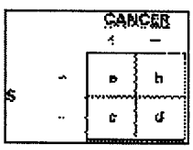

Figure 5. Table for determining the association between the surrogate endpoint (S) and cancer. Relative risk is determined by [a/(a+b)]/[c/(c+d)]. The odds ratio is determined by ad/bc.

ISSUE 1: SURROGATE VERSUS CANCER

For a marker to be a reasonable surrogate for a given cancer, it must have some relation to that cancer. In evaluating a potential surrogate, therefore, there must be some way of establishing the connection between the marker and malignancy.

Ecologic studies may provide useful, if indirect, information on this connection. Studies are considered to be *ecologic,* or aggregate, in that individual-level information is not used; rather, an average marker value is obtained from a sample taken from specific populations (e.g., Seventh Day Adventists versus non-Adventists), and that marker value is then related to the overall risk of cancer in those populations. Several studies, for example, have compared mean proliferation indices in groups at varying risk of cancer. [22] In such studies, however, one cannot be certain that those who are marker-positive are the ones with increased incidence of cancer.

This ecologic problem is obviated by moving to individual-level observational epidemiologic studies, whether case-control or cohort. Such studies are important tools for examining the relation between a putative surrogate and cancer. Blood estrogen levels have been shown in several studies to be directly associated with breast cancer, a relation that had to be established before estrogens could be considered a surrogate for breast malignancy. [36] Human papillomavirus infection, a potential surrogate for cervical cancer, has been shown to be highly associated with risk of severe cervical neoplasia. [34] Observational studies may also be nested in trials. In trials with adenomatous polyp formation as the primary endpoint, for example, it is possible to examine the relation of colorectal epithelial cell proliferation measures to subsequent adenoma recurrence. In one such study, Baron et al found no relation between calcium carbonate supplementation and epithelial cell proliferation measured 1 year later, [2] even though calcium did reduce overall adenoma recurrence. [1] This finding certainly suggests that proliferation measures are problematic surrogates for colorectal neoplasia and cancer in studies with calcium supplements as the

main intervention/exposure.

(Note: References here are to studies with adenomas or cervical intraepithelial neoplasia (CIN) endpoints. Although these are only neoplastic cancer precursors, they are here, for purposes of discussion, considered as proxies for cancer. As I discuss later, the validity of these precursor endpoints is not ironclad.)

The epidemiologic parameter AP may be useful in determining the importance of alternative pathways and thereby evaluating the relation between the potential surrogate and cancer. In the simple linear causal model of Figure 1 , the AP for the surrogate is 1.0. When at least one pathway exists that is alternative to the pathway containing the surrogate, as in Figure 2 , the AP for the surrogate is less than 1.0. If, although it is less than 1.0, the AP is still relatively high, that suggests that the alternative pathway M2 plays a small role in the genesis of the cancer. An AP substantially lower than 1.0 for the surrogate implies that one or more alternative pathways is indeed operative.

ISSUE 2: INTERVENTION (OR EXPOSURE) VERSUS SURROGATE

For a given potential surrogate marker to be valid with respect to a particular intervention (or exposure), there must be some relation between the intervention (or exposure) and the marker. In an experimental setting, it must be shown that the intervention changes the marker. In an observational context, an association must be observed between an exposure and marker positivity.

This question can be addressed in relatively small metabolic intervention studies. Several studies, for example, have examined the effect of dietary change or supplementation on epithelial cell proliferation [20] ; others have investigated the effect of fat modification [28] or alcohol consumption [29] on blood or urine estrogen levels. This question can also be examined in a case-control or cohort study. Schiffman et al, for example, showed a strong association between reproductive risk factors, particularly the number of sexual partners, and HPV infection. [34] Investigators will be able to look at the relation between breast cancer risk factors and serum hormone levels. Ecologic studies may also provide indirect information. One could, for example, examine the mean colorectal epithelial cell proliferation index in populations with different average consumption of dietary fat.

ISSUE 3: INTERVENTION (EXPOSURE) VERSUS SURROGATE VERSUS CANCER

Having investigated issues 1 and 2, the investigator may have determined that (1) the potential surrogate is probably causally linked to cancer, and (2) the surrogate is indeed linked to a given intervention or exposure. Suppose, however, with reference to issue 1, that the AP is less than 1.0, indicating that alternative pathways bypassing the surrogate are operative in carcinogenesis. It would still be valuable to ascertain the relative importance of the intervention (or exposure) → surrogate → cancer pathway, as opposed to the alternative intervention (or exposure) → other markers → cancer pathways. The relative importance can be determined by examining the extent to which the putative surrogate mediates the relation between the intervention (or exposure) and cancer. In other words, the extent to which surrogate endpoint status accounts for any observed intervention effect or exposure association can be determined. To carry out such analyses, one must integrate an assay for the surrogate into either clinical trials or

observational epidemiologic studies with information about both the intervention (or exposure) and the cancer (or severe neoplasia).

As an example, investigators have used a case-control study to look at the extent to which HPV infection mediated the association between the number of sexual partners and dysplasia. [35] As Table 1 shows, the number of sexual partners and cervical dysplasia risk were strongly and directly associated. When the relation between number of sexual partners and dysplasia was adjusted for the presence or absence of HPV infection, the relative risk for number of sexual partners dropped dramatically. This finding suggests that most of the association between number of partners and cervical dysplasia is caused by HPV infection.

TABLE 1 -- CERVICAL DYSPLASIA ODDS RATIO FOR NUMBER OF SEXUAL PARTNERS, UNADJUSTED AND ADJUSTED FOR HPV STATUS

	Number of Sexual Partners				
Odds Ratio	*1*	*2*	*3- 5*	*6- 9*	*10+*
Unadjusted	1.0	1.7	3.1 *	4.7 *	4.4 *
Adjusted for HPV status	1.0	1.0	1.1	1.5	1.6

HPV = human papillomavirus

* $P<0.05$

Researchers obtaining blood specimens from participants in large cohort studies will be able to investigate whether serum hormone levels mediate the relation between reproductive risk factors and breast cancer. A dietary modification or dietary supplement study of colorectal neoplasia, from which rectal biopsy specimens are obtained for mucosal proliferation assays, can provide information on the extent to which any observed effect from diet or supplement is attributable to proliferation changes.

The statistical aspects of mediation analysis are important and are an area of current research. [7] [15] One can investigate whether a potential surrogate marker mediates a relation between an intervention (or exposure) and cancer by using stratified analyses or by multiple regression techniques. Generally, the greater the intervention effect or exposure association, the fewer study participants are needed in a mediation analysis. Because exposure relative risks tend to be considerably larger than the intervention effects observed in experimental epidemiologic studies (trials), mediation analyses may be more likely to provide interpretable data in the observational epidemiologic setting. Specific genetic mutations or polymorphisms, if they are shown to yield high relative risks for specific cancers, may prove a valuable source of mediation analyses of various potential biochemical or cellular surrogates.

Mediation analyses, in which the intervention or exposure relative risks are adjusted for the value of the potential surrogate marker, may yield null results. Such null findings suggest that the potential surrogate does not mediate the relation between intervention (or exposure) and cancer. Even with null results, however, there are two possible scenarios under which the surrogate could still be on the causal pathway to cancer. The first, illustrated in Figure 2 , occurs when there is an alternative pathway from an exposure E1 to cancer through a second marker. In other words, the surrogate is not a necessary step between E1 and cancer. The degree to which the exposure E1-cancer relation is attenuated after adjustment for the surrogate depends on the relative contributions of the alternative pathways--one

through the surrogate, the other through another marker--to the development of cancer. These relative contributions are probably unknown.

The second scenario is illustrated in Figure 6 . Some unknown factor leads to surrogate marker positivity. The positive surrogate marker also requires the exposure E1 as a cofactor for the development of cancer. Thus, the surrogate is on the pathway to cancer, but adjustment for surrogate marker status will not necessarily reduce the relative risk of the exposure to 1.0. The surrogate does not mediate the known risk factor but does mediate the unknown risk factor.

Figure 6. Some unknown factor leads to surrogate marker positivity, and the positive surrogate marker requires the exposure E1 as a cofactor for the development of cancer. The surrogate is on the pathway to cancer, but adjustment for surrogate marker status will not necessarily reduce the relative risk of the exposure to 1.0. The surrogate mediates the unknown risk factor but not the known one.

The presence of interaction in mediation analyses is also a consideration. In Figure 3 , an intervention affects both the surrogate and another **intermediate** marker, M2. Theoretically, it is possible that the intervention can affect the surrogate and the other marker in offsetting ways. In that instance, the mediation analysis will demonstrate a significant interaction between the intervention and the surrogate marker; that is, the cancer rate among surrogate-positive participants will differ according to whether they are in the intervention (exposed) or control (unexposed) group. Such an interaction indicates that the surrogate does not fully mediate the intervention effect. The surrogate, however, does indeed lie on a single dominant causal pathway.

SURROGATES WITH HIGH LIKELIHOOD OF VALIDITY

Unlike putative surrogates such as epithelial cell proliferation or blood hormone levels, for which, as discussed previously, validity is problematic, considerable evidence supports the relative validity of a few *downstream* surrogate markers (that is, those close to cancer on the causal pathway).

An overwhelmingly large proportion of cervical cancer requires prior, persistent HPV infection. (Some immunologic deficit or nutritional or environmental cofactor may be involved in the development of persistent HPV infection.) Persistent HPV infection results in inactivation, by the E6 and E7 proteins of the HPV genome, of *p53* and *pRb* tumor suppressor genes, leading in turn to increasingly severe intraepithelial neoplasia and, eventually, to cancer. Only a very small proportion of cervical cancer, at most, arises from inactivation of tumor suppressor gene products occurring by mutation in the absence of HPV infection. Because most cervical cancer does occur through persistent HPV infection, an intervention that eliminates or reduces such infection would have a high likelihood of decreasing the incidence of cervical cancer.

With regard to cervical cancer, it should be mentioned that CIN, especially CIN3, is considered a strong surrogate for cancer and has been used as an endpoint in a number of epidemiologic studies. A very high percentage of CIN3 will progress to cancer in 20 years; only a very small fraction regresses. In fact, CIN3 is very close to being invasive cancer.)

A second example is Barrett's esophagus, a metaplastic change from squamous to columnar epithelium in the lower esophagus that is thought to be a necessary precursor to most cases of esophageal adenocarcinoma. [18] Gastric acid reflux is regarded as the primary precipitant of this metaplastic change; other factors may operate in the transition from Barrett's epithelium through dysplasia to

adenocarcinoma. As in the example of cervical cancer and persistent HPV infection, a small proportion of esophageal adenocarcinomas seem to arise from esophageal submucosal glands, independent of the Barrett's epithelium pathway. Nevertheless, an intervention (e.g., photoablation or electrocoagulation [5]) that eradicates the Barrett's epithelium would probably greatly reduce the incidence of esophageal adenocarcinoma. The operative word here is "eradication." Some have argued that photoablation causes epithelialization over remaining "nests" of Barret's cells. If that were the case, the patient's condition would been misclassified (the observed eradication would cause the individual to be classified as "marker negative," when in fact the true state is "marker positive").

THE ADENOMATOUS POLYP AS SURROGATE ENDPOINT

The marker that has received perhaps the greatest attention as a potential surrogate endpoint in clinical trials--one for which inferences to cancer are considered to be strong--is the adenomatous polyp (adenoma). Colorectal adenomas are attractive candidates for cancer surrogacy in research studies because of their high prevalence (making it possible to recruit study participants) and high recurrence rate (at about 10% or more per year, nearly two orders of magnitude greater than the incidence of cancer). The underlying biologic rationale for the use of adenoma endpoints in epidemiologic studies and clinical trials is the strong pathologic, cell biologic, and molecular biologic evidence showing a relation between this marker and colorectal cancer. This adenoma-to-carcinoma sequence is supported by studies demonstrating carcinomatous foci in adenomas and adenomatous foci within carcinomas, by experiments showing the malignant transformation of adenoma cell lines, and by studies identifying common mutations in adenomatous and carcinomatous tissue. [31] It is widely acknowledged that the alternative pathway to colorectal cancer, through flat dysplasia rather than the raised polypoid neoplasm, is a minor pathway for most people with intact epithelia (those, for example, without ulcerative colitis). An intervention reducing the recurrence of adenomas in the large bowel would therefore probably decrease the incidence of colorectal cancer, thus making adenoma recurrence a reasonably valid surrogate marker.

It is worth noting, however, that even the adenoma is not an ironclad surrogate, and some inferential difficulties remain with studies of adenoma recurrence. Recurrent adenomas represent a spectrum of neoplastic changes from normal mucosa through the development of a small adenoma. The results of adenoma recurrence trials may be misleading if the intervention factor operates later in the neoplastic process, that is, during the growth of a small adenoma into a large adenoma or the transformation of a large adenoma to cancer. A false null result for recurrent adenomas may result if the intervention operates only in the later stages of neoplasia. A positive result, however, suggests that cancer would be reduced, because large adenomas and cancers derive from small adenomas.

A second inferential difficulty with adenoma recurrence as a surrogate endpoint arises from the probable biologic heterogeneity of adenomas. Only a relatively small proportion of adenomas progress to cancer. Suppose that one type, the "bad" adenoma that progresses to cancer, is caused by exposures E1 and E2, as in Figure 7 . The second type, the "innocent" adenoma, is caused by the same exposure (E1) but in concert with exposure E3. Imagine an intervention that works only on exposure E3. The pool of innocent adenomas could be reduced, thereby yielding a statistically significant reduction in adenoma formation in the trial, but in fact the incidence of bad adenomas and cancer would be unaffected. This process could work the other way as well: at most, a small reduction in all adenomas might be observed (the bad ones being only a small proportion of all adenomas), even though the intervention truly decreases the formation of bad adenomas and, therefore, reduces the incidence of cancer.

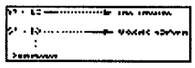

Figure 7. Exposures E1 and E2 causes the "bad" adenoma that progresses to cancer. The "innocent" adenoma is caused by the same exposure E1, but in concert with E3. An intervention that affected only exposure E3 would result in reduced "innocent" adenoma formation, but would leave the "bad" adenoma (and, by inference, cancer) unaffected.

MUST A MARKER BE A NECESSARY STEP ON THE PATHWAY TO CANCER TO BE A VALID SURROGATE?

Markers such as persistent HPV infection and adenoma formation are close to being necessary steps on the pathways to cervical and colorectal cancer, respectively. In other words, alternative pathways bypassing these markers are minor at best. Therefore, it is unlikely that the action of an intervention or exposure through an alternative pathway can offset the action through the surrogate's pathway. When a marker is a necessary, or almost necessary step on the causal pathway to cancer, there is prior confidence that the marker is a reasonably valid surrogate. (Other inferential difficulties can arise, however, such as those discussed previously for adenoma recurrence).

That is not to say, however, that to be a valid surrogate a marker must be a necessary step toward cancer, with an AP near 1.0. Even when alternative pathways bypassing the surrogate do exist, they do not necessarily offset action through the surrogate marker. In other words, going back to cell proliferation as an endpoint, even if there were an alternative pathway through apoptosis, a given intervention agent such as aspirin or calcium might not affect apoptosis in any substantial way. The effect of the agent on proliferation, therefore, would be translated into some reduction in cancer incidence. [32] The problem is that the possible offsetting pathways are not known in advance; the marker must be tested in large-scale epidemiologic studies or trials.

The virtue of the necessary or close-to-necessary markers such as HPV infection or adenoma is that the offsetting alternative pathways are of little significance. The causal logic, even without empiric testing in large-scale studies, dictates confidence in these markers.

IS A SURROGATE VALID FOR ONE INTERVENTION VALID FOR ANOTHER?

Figure 8 reprises the simple idealized scheme from Figure 1 but adds another exposure, E2. (As before, the exposure here can refer to an intervention agent or to a risk factor.) In Figure 8 , both E1 and E2 operate through a single surrogate on the path to cancer. Because in this scenario the surrogate is a necessary component of the cancer pathway, the validity of this surrogate is exposure-independent. In other words, any other exposure, E2, that affects cancer must operate through the surrogate. The surrogate is valid for studies of E2 as well as those of E1.

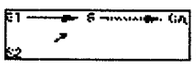

Figure 8. Both E1 and E2 operate through a single surrogate on the path to cancer. The surrogate is a necessary component of the cancer pathway, so the validity of the surrogate is exposure-independent. The surrogate is valid for studies of E1 and E2.

In Figure 9 , with E2 entering into the more complex scenario depicted in Figure 2 , the existence of a nontrivial alternative pathway through M2 means that the validity of the surrogate S may be exposure-dependent. Even if E1 works primarily through the surrogate and affects M2 minimally, suggesting that

the surrogate is reasonably valid for E1-cancer studies, one cannot assume that the E2-M2-cancer pathway plays a similarly minor role in carcinogenesis. A given agent, for example, might influence colorectal carcinogenesis largely through its influence on cell proliferation (see Fig. 4) . In this scenario, cell proliferation is a probable valid surrogate for colorectal cancer. A second agent, however, might have minimal effect on cell proliferation but could increase apoptosis (or activate some other comparable alternative pathway) sufficiently to decrease cancer incidence. Focusing only on cell proliferation would give a falsely pessimistic impression of the efficacy of the second agent. In other words, even if proliferation markers are established as valid surrogates for a colorectal cancer in calcium studies, one cannot be certain that this marker is a valid surrogate for vitamin E or aspirin studies. The marker may be valid for these other agents, but the existence of alternative pathways engenders uncertainty that could only be avoided by specific validation studies for these other agents.

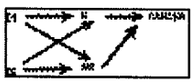

Figure 9. The existence of a nontrivial alternative pathway (through M2) means that the validity of the surrogate S may be exposure-dependent. Even if E1 works primarily through the surrogate and affects M2 minimally, suggesting that the surrogate is reasonably valid for E1 cancer studies, it cannot be assumed that the E2-M2-cancer pathway plays a similarly minor role in carcinogenesis.

INTERPRETING POTENTIAL SURROGATE MARKER DATA: EPIDEMIOLOGIC AND STATISTICAL CONSIDERATIONS

Common sense and judgment are essential in any epidemiologic study or clinical trial involving potential surrogate markers. The traditional epidemiologic causal criteria can be applied to data from these studies involving surrogates. In considering persistent HPV infection as a possible surrogate, are the results biologically plausible? Yes, there is good reason to think that the strong epidemiologic association between the number of sexual partners and cervical neoplasia or cancer is explained by the distribution of HPV infection. Are data from multiple studies consistent? Several studies have now demonstrated the relation between HPV and cervical neoplasia or cancer, and at least one has shown that HPV mediates the association between reproductive risk factors and cervical neoplasia. Are the measures of effect (the RR, AP) strong? They are for the connection between HPV and cervical neoplasia or cancer. Is the reduction of relative risk in the mediation analysis substantial? It was in the reproductive risk factors-HPV-cervical neoplasia mediation analysis. The totality of evidence suggests that HPV infection is a reasonably valid surrogate for cervical neoplasia or cancer.

All biomarkers are measured with some error. Two important statistical issues need to be considered. First, a potential surrogate is useful (and ultimately valid) only if it can discriminate among study participants, those in the different treatment arms of a trial or the various exposure categories in an epidemiologic study. Discrimination is possible only if the interparticipant variability in the surrogate values is greater than the intraindividual variability. (Intraindividual variability may arise, for example, from differences in marker values obtained from different tissue areas, measured at different time points, or read by multiple readers.) The intraclass correlation coefficient (ICC), the proportion of all variability (inter- plus intraparticipant variability) caused by interparticipant variability, must be reasonably large. [13] (ICC=interparticipant variability/[interparticipant variability+intraparticipant variability]).

Intraparticipant variability may be reduced (and the ICC thereby increased) by taking repeat samples, such as multiple biopsies from different areas or multiple blood draws over time. At a minimum, therefore, data are required on the potential surrogate marker's components of variance to establish the minimum number of marker samples needed for meaningful discrimination among study participants. In the absence of such data, one cannot be certain that null findings for a potential surrogate reflect a true

lack of effect or association or simply the attenuating influence of random sources of intraindividual variation.

These data have not been routinely collected in marker studies. Few studies have provided data on potential surrogate marker variability, particularly with respect to time-to-time variability. A notable exception are recent investigations attempting to estimate the number of estradiol measurements necessary to discriminate reasonably among individuals. [19] Studies of colorectal epithelial cell proliferation are underway. [23] Quality-control studies designed to obtain data on the variability characteristics of potential surrogate markers are essential.

Second, even if the ICC is acceptable, measurement error will tend to attenuate findings from each of the three studies previously discussed. The intervention (exposure)-marker and marker-cancer associations will be attenuated by error in marker measurement. In mediation analyses, the expected attenuation of the intervention effect (exposure association) will itself be attenuated, so the marker-adjusted effect (association) will be inflated.

ADDITIONAL ISSUES IN EVALUATING POTENTIAL SURROGATE MARKERS

Comparing one potential surrogate marker assay with another (e.g., comparing proliferating cell nuclear antigen with bromodeoxyuridine or tritiated thymidine cell proliferation assays [11]) can provide valuable information on assay characteristics and on the biology of the phenomenon under study. Such a comparison, however, does not constitute surrogate validation. The close association of a newer marker with an older one does not in itself overcome the inferential limitations of the older marker.

A marker not directly on the causal pathway to cancer may still be closely linked with a component of that pathway so that it does constitute a reasonable surrogate marker. One possible example is the presence of micronuclei, which have been detected in epithelial cells from oral, esophageal, bronchial, and colorectal tissue. [16] Many micronucleated cells are nonviable and therefore cannot be a direct cellular precursor of a malignant tumor. The overall prevalence of micronucleated cells, however, might strongly reflect genetic damage in other cells that do eventually undergo malignant transformation and clonal expansion.

An increasingly popular (if not explicit) approach to cancer research involves a two-stage strategy. In one set of studies, the relation between an intervention agent or exposure and a potential surrogate marker is examined. In a separate set of studies, the relation between the marker and the cancer is investigated. For example, investigators have looked at (1) the effect of alcohol consumption on blood estrogen levels in metabolic studies, and (2) the association between blood estrogen levels and breast cancer in prospective cohort studies. Strong links for the marker in both sets of studies would give at least some evidence that the marker is a valid surrogate for cancer. This two-stage evidence, however, is less than absolute for surrogate marker validation, because the intervention or exposure may be related to a second marker in a way that offsets the effect through the first marker (see Figs. 2 and 9) . This potential offsetting phenomenon cannot be detected in the two-stage separate studies.

Surrogate markers should also be considered in a broader context of multiple disease endpoints and of intervention or treatment toxicity. A surrogate marker might give the "right" answer about cancer for a given intervention but nevertheless give little or no information about important adverse events that bear heavily on any overall evaluation of the intervention. Suppose, for example, that there is a valid tissue or blood marker for breast cancer, one that gives the right answer about a promising hormone-modulating

intervention. That breast cancer surrogate will tell nothing about the potential of the intervention to increase the incidence of stroke. One could examine potential surrogates for stroke, but uncertainties remain about the relation of the breast cancer surrogate and the actual occurrence of a serious cerebrovascular event. In the absence of clinical-trial evidence with explicit stroke endpoints, could cholesterol determinations or platelet aggregation studies be accepted as definitive proof that the theoretical intervention does not raise the risk of cerebrovascular catastrophe? The fairly obvious negative answer to this question illustrates still another dimension of difficulty arising from the exclusive reliance on surrogate marker studies.

SUMMARY

Because studies with surrogate cancer endpoints can be smaller, faster, and substantially less expensive than those with frank cancer outcomes, the use of surrogate endpoints is undeniably attractive. This attractiveness is likely to grow in coming years as the rapidly advancing discoveries in cell and molecular biology generate new therapies requiring testing and new markers that could plausibly serve as surrogates for cancer.

Surrogate endpoint studies can certainly be suggestive. They continue to play a legitimate role in phase II studies, and they may give the right answers about intervention effects on or exposure associations with cancer.

The problem is the uncertainty attached to most potential surrogates. Except for those few surrogates that are both necessary for and developmentally relatively close to cancer, the existence of plausible alternative pathways makes inferences about cancer from many surrogates problematic. Merely being on the causal pathway to cancer does not in itself constitute surrogate validity. It is the totality of causal connections that is critical. There is, unfortunately, a fairly extensive history of quite plausible surrogate markers giving the wrong answer about various chronic disease therapies. [14] There is no reason to believe that cancer surrogacy is immune to such inferential difficulties.

This article is, in part, an invitation, even a plea, for researchers to carry out the investigations necessary to evaluate potential surrogates, particularly surrogate-cancer studies and intervention or exposure-surrogate-cancer mediation analyses. Such studies are needed to generalize from surrogate endpoint findings to cancer. There is, however, an implicit and perhaps unavoidable irony here: the large, long, expensive studies required to evaluate potential surrogates fully are precisely the studies that surrogates were designed to replace. The exposure dependence alluded to earlier complicates matters further: establishing validity for a given surrogate for one intervention or exposure vis-a-vis cancer does not necessarily translate into validity for another intervention or exposure.

One can enhance the inferential strength of surrogacy by using further "downstream" markers. Results of trials with CIN3 as an endpoint are arguably more persuasive than those from intervention studies with HPV infection endpoints. Similarly, one could consider only the advanced adenoma ($\geq$ 1 cm, villous elements, or high-grade dysplasia) as the primary endpoint in adenoma recurrence trials. The inferential gain, however, comes with substantial costs: studies with CIN3 endpoints must be much larger than those with HPV infection endpoints; adenoma recurrence trials with sufficient rates of recurrence of advanced adenomas must be five or six times larger than trials with any recurrent adenomas as endpoints. A law emerges here: in using surrogate endpoints, inferential certainty is directly associated with study cost. In other words, one gets what one pays for.

The problems inherent in using surrogate endpoints need not be regarded as a cause for pessimism in cancer research. If anything, the limitations of surrogacy are reminders of the complexity of cancer causation and affirm the continued importance of large clinical trials and observational epidemiologic studies with explicit cancer endpoints.

References

1. Baron JA, Beach M, Mandel JS, et al: Calcium supplements for the prevention of colorectal adenomas. New Engl J Med 340:101-107, 1999 Abstract

2. Baron JA, Tosteson TD, Wargovich MJ, et al: Calcium supplementation and rectal mucosal proliferation: A randomized controlled trial. J Natl Cancer Inst 87:1303-1307, 1995 Abstract

3. Baron JA, Wargovich MJ, Tosteson TD, et al: Epidemiological use of rectal proliferation measures. Cancer Epidemiol Biomarker Prev 4:57-61, 1995

4. Bedi A, Pasrich PJ, Akhtar AJ, et al: Inhibition of apoptosis during development of colorectal cancer. Cancer Res 55:1811-1816, 1995 Abstract

5. Berenson MM, Johnson TD, Markowitz NR, et al: Restoration of squamous mucosa after ablation of Barrett's esophageal epithelium. Gastroenterology 104:1686-1691, 1993 Abstract

6. Blattner WA: Retroviruses. *In* Evans AS (ed): Viral Infections in Humans, ed 3. New York, Plenum Medical Book Co, 1989, pp 545-592

7. Buyse M, Molenberghs G: Criteria for the validation of surrogate endpoints in randomized experiments. Biometrics 54:1014-1029, 1998 Abstract

8. Counts JL, Goodman JI: Alterations in DNA methylation may play a variety of roles in carcinogenesis. Cell 83:13-15, 1995 Citation

9. Dawsey SM, Fleischer DE, Wang GQ, et al: Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. Cancer 83:220-231, 1998 Abstract

10. Dorgan JF, Longcope C, Stephenson HE, et al: Relations of prediagnostic serum estrogen and androgen levels to breast cancer risk. Cancer Epidemiol Biomarker Prev 5:533-539, 1996

11. Einspahr J. Alberts D, Xie Tailiang, et al: Comparison of proliferating cell nuclear antigen versus the more standard measures of rectal mucosal proliferation rates in subjects with a history of colorectal cancer and normal age-matched controls. Cancer Epidemiol Biomarker Prev 4:359-366, 1995

12. Fearon ER: Genetic alterations underlying colorectal tumorigenesis. Cancer Surveys 12:119-136, 1992

13. Fleiss JL: The Design and Analysis of Clinical Experiments. New York, John Wiley & Sons, 1986, pp 1-5

14. Fleming TR, DeMets DL: Surrogate end points in clinical trials: Are we being misled? Ann Intern Med 125:605-613, 1996 Citation

15. Freedman LS, Graubard BI, **Schatzkin** A: Statistical validation of **intermediate** endpoints for chronic diseases. Stat Med 11:167-178, 1992 Abstract

16. Garewal HS, Ramsey L, Kaugars G, et al: Clinical experience with the micronucleus assay. J Cell Biochem (suppl)

17F:206-212, 1993

17. Groopman JD, Wogan GN, Roebuck BD, et al: Molecular biomarkers for aflatoxins and their application to human cancer prevention. Cancer Res 54 (suppl):1907s-1911s, 1994 Abstract

18. Haggitt RC: Barrett's esophagus, dysplasia, and adenocarcinoma. Hum Pathol 25:982-993, 1994 Abstract

19. Hankinson SE, Manson JE, Spiegelman D, et al: Reproducibility of plasma hormone levels in postmenopausal women over a 2-3-year period. Cancer Epidemiol Biomarker Prev 4:649-654, 1995

20. Holt PR, Atillasoy EO, Gilman J, et al: Modulation of abnormal colonic epithelial cell proliferation by low-fat dairy foods: A randomized, controlled trial. JAMA 280:1074-1079, 1998 Abstract

21. Karlan BY: Screening for ovarian cancer: What are the optimal surrogate endpoints for clinical trials? J Cell Biochem 23 (suppl):227-232, 1995

22. Lipkin M, Blattner WA, Gardner EJ, et al: Classification and risk assessment of individuals with familial polyposis, Gardner's syndrome, and familial non-polyposis colon cancer from [^{3} H] thymidine labeling patterns in colonic epithelial cells. Cancer Res 44:4201-4207, 1984 Abstract

23. Lyles CM, Sandler RS, Keku TO, et al: Reproducibility and variability of the rectal mucosal proliferation index using proliferating cell nuclear antigen immunohistochemistry. Cancer Epidemiol Biomarker Prev 3:597-605, 1994

24. Misset JL, Mathe G, Santelli G, et al: Regression of bronchial epidermoid metaplasia in heavy smokers with etretinate treatment. Cancer Detect Prev 9:167-170, 1986 Abstract

25. Mitchell MF, Hittelman WN, Hong WK, et al: The natural history of cervical intraepithelial neoplasia: An argument for **intermediate** endpoint biomarkers. Cancer Epidemiol Biomarker Prev 3:619-626, 1994

26. Munoz N: Is *Helicobacter pylori* a cause of gastric cancer? An appraisal of the seroepidemiological evidence. Cancer Epidemiol Biomarker Prev 3:445-451, 1994

27. Prentice RL: Surrogate endpoints in clinical trials: Definition and operational criteria. Stat Med 8:431-440, 1989 Abstract

28. Prentice R, Thompson D, Clifford C, et al: Dietary fat reduction and plasma estradiol concentration in healthy premenopausal women. J Natl Cancer Inst 82:129-134, 1990 Citation

29. Reichman ME, Judd JT, Longcope C, et al: Effects of moderate alcohol consumption on plasma and urinary hormone concentrations in premenopausal women. J Natl Cancer Inst 85:722-727, 1993 Abstract

29A. Ries LAG, Kosary CL, Hankey BF, et al (eds): SEER Cancer Statistics Review, 1973-1996. Bethesda, MD, National Cancer Institute, 1999

30. Saftlas AF, Wolfe JN, Hoover RN, et al: Mammographic parenchymal patterns as indicators of breast cancer risk. Am J Epidemiol 129:518-526, 1989 Abstract

31. **Schatzkin** A, Freedman LS, Dawsey SM, et al: Interpreting precursor studies: What polyp trials tell us about large bowel cancer. J Natl Cancer Inst 86:1053-1057, 1994 Citation

32. **Schatzkin** A, Freedman LS, Schiffman MH, et al: The validation of **intermediate** endpoints in cancer research. J Natl Cancer Inst 82:1746-1752, 1990 Abstract

33. Schiffman MH: Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia. J

Natl Cancer Inst 84:394-398, 1992 Citation

34. Schiffman MH, Bauer HM, Hoover RN, et al: Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J Natl Cancer Inst 85:958-964, 1993 Abstract

35. Schiffman MH, **Schatzkin** A: Test reliability is critically important to molecular epidemiology: An example from studies of human papillomavirus infection and cervical neoplasia. Cancer Res 54:1944s-1947s, 1994 Abstract

36. Toniolo PG, Levitz M, Zeleniuch-Jacquotte A, et al: A prospective study of endogenous estrogens and breast cancer in postmenopausal women. J Natl Cancer Inst 87:190-197, 1995 Abstract

37. Tsiatis AA, DeGruttola V, Wulfsohn MS: Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. J Am Stat Assoc 90:27-37, 1995

38. Wargovich MJ: Precancer markers and prediction of tumorigenesis. *In* Young GP, Rozen P, Levin B (eds): Prevention and Early Detection of Colorectal Cancer. London, WB Saunders, 1996, pp 89-101

39. Women's Health Initiative Study Group: Design of the Women's Health Initiative Clinical Trial and Observational Study. Controlled Clinical Trials 19:61-109, 1998